Time Series Project

Name: Lei Cheng

Email: thatstheway@126.com

Session: Spring 2015

Introduction

Gold has both the goods attribute and the monetary attribute. The gold price is the focus of attention both at home and aboard. The gold price is influenced by many factors and experienced the different degree of rise and fall. Since gold investment becomes the hot issue, the future of gold price analysis is important. In this paper, we will do some research on the historic gold price and the aim is to establish a time series model for the gold price.

Data

The analysis is based on the monthly gold price from year 2009 to year 2015, 76 points are picked up to do our analysis.

Four types of data will be used in the analysis:

1. Initial data for the gold price
2. First difference of the data for the gold price
3. Log-transformation of the data for the gold price
4. First difference of the log-transformation for the gold price

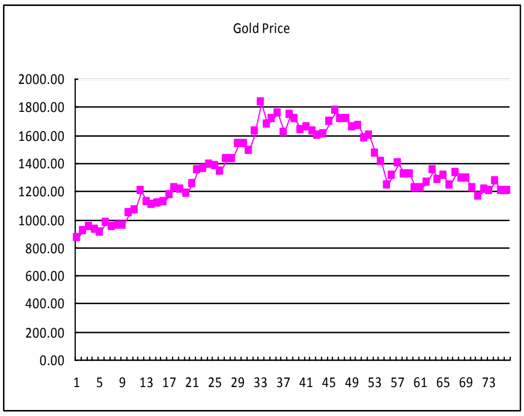
The above figures are shown in appendix.

Methodology

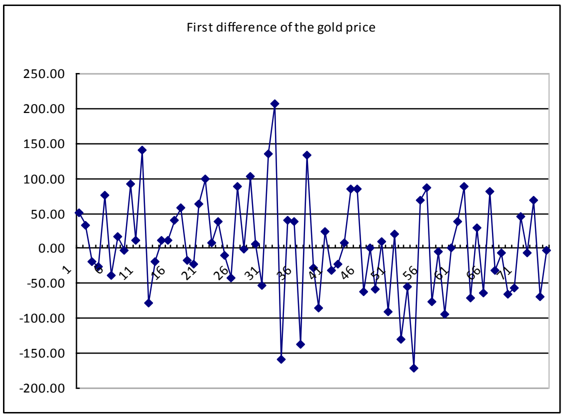
The data graph and the sample autocorrelation function will be used in this paper to do research on the stationary, seasonality and periodicity of the data. Furthermore, ARIMA (p, d, q) model will be used to matching the data. Finally, the most appropriate model will be picked up based on the linear regression and Shapiro-Wilk normality test results. R software will be used for the statistical computing.

Analysis

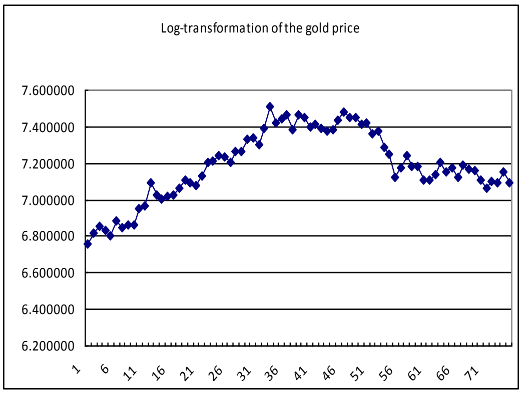
First, we use Excel line chart to draw the four types of data as followed:

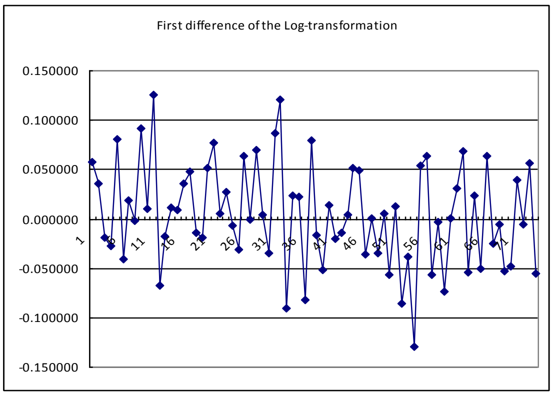


1. Initial data for the gold price



2. First difference of the data for the gold price

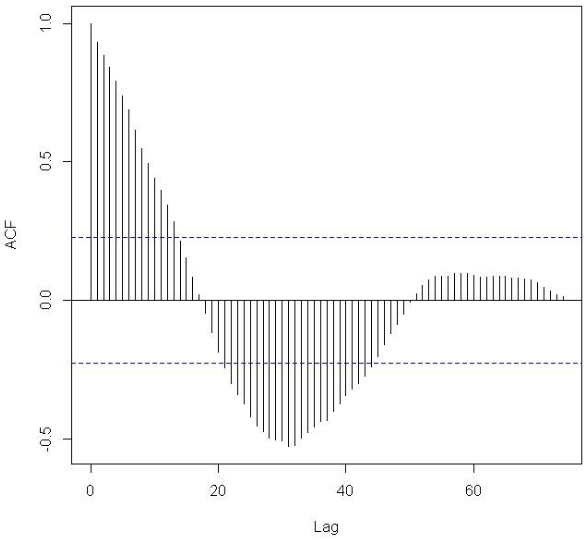


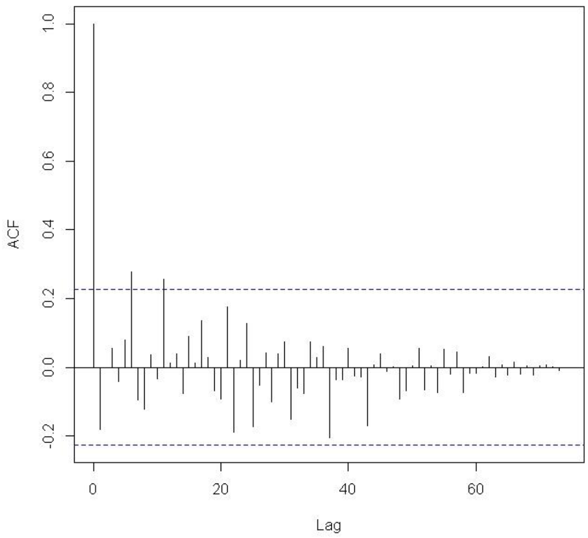


3. Log-transformation of the gold price

4. First difference of the log-transformation

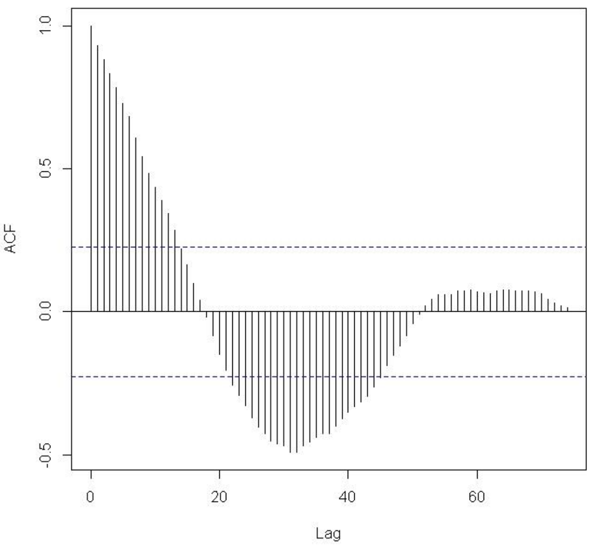
Second, we use R software to draw ACF Diagram for the four types of data as followed:

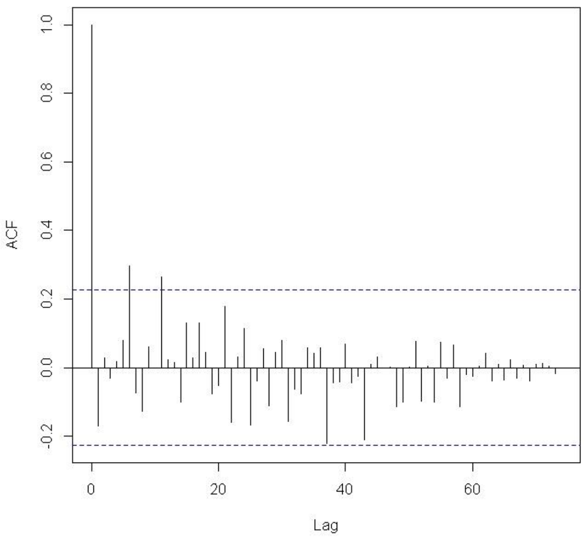




1. Initial data for the gold price

2. First difference of the data for the gold price





3. Log-transformation of the gold price

4. First difference of the log-transformation

We know that in the simple autocorrelation function diagram, more points lie in the confidence region, more likely a white noise is successfully modelled. From the graph above, the advantage for this character is not very obviously to pick up the data directly. Therefore, next we will try to apply the ARIMA model on the four types of data and finally find out the best one. In detail, ARIMA(p,d,q) model with p, d and q lower than 2 will be used to do the regression according to the principle of parsimony.

We will check six models one by one as followed:

1. ARIMA(0,0,0): White noise on the gold price

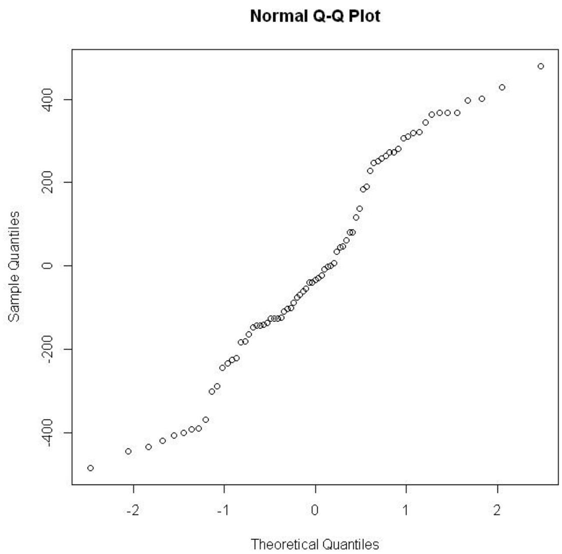
   The formula is

   $Y_t = \mu + e_t$

   Where μ is the mean value of the data and $\mu = 1348.933$

   Therefore, the formula can be written as $Y_t = 1348.933 + e_t$

   We use qqnorm function in R software to get the q-q plot for the residuals of the model as followed:



   The Shapiro-Wilk normality test result shows that the statistic value is 0.9625 and the p-value is 0.0256. The p-value is lower than 0.05 and the residuals are not symmetrical, therefore, the model is rejected.

2. ARIMA(1,0,0) on the gold price

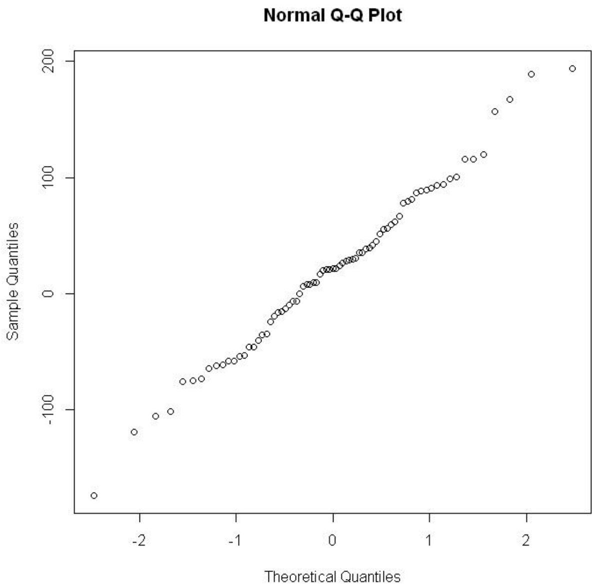
   The formula is

   $Y_t = \mu + \phi_1 Y_{t-1} + e_t$

   Where μ is the mean value of the data, ϕ_1 is the least squares estimation.

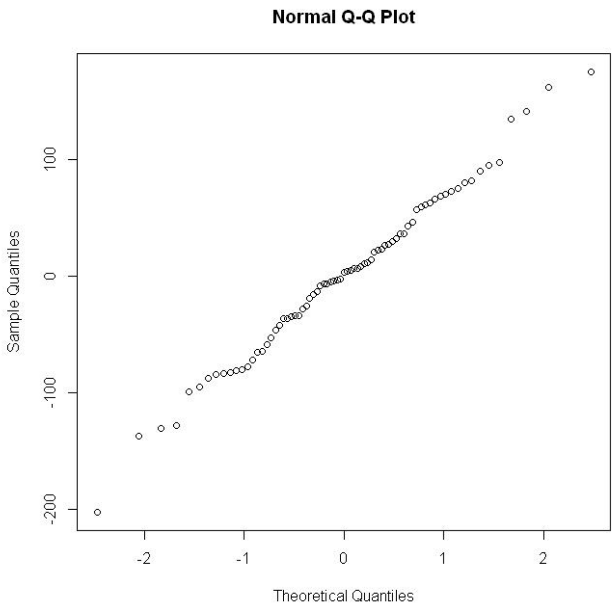
   Therefore, the formula can be written as $Y_t = 1348.933 - 0.9818Y_{t-1} + e_t$

   We get the q-q plot for the residuals of the model as followed:



The Shapiro-Wilk normality test result shows that the statistic value is 0.9913 and the p-value is 0.8907. The p-value is much higher than 0.05 and the residuals are symmetrical. Next we will check the rest of models to find out if there exists a better one than this model.

3. ARIMA(0,1,0) on the gold price
The formula is
$W_t = Y_t - Y_{t-1} = e_t$
We get the q-q plot for the residuals of the model as followed:



The Shapiro-Wilk normality test result shows that the statistic value is 0.9923 and the p-value is 0.9329. The p-value and the q-q plot are satisfied the request and better than the second model. Therefore we still need to check the rest of models to find out if there exists a better one.

4. ARIMA(0,0,0): White noise on the log-transformation of the gold price

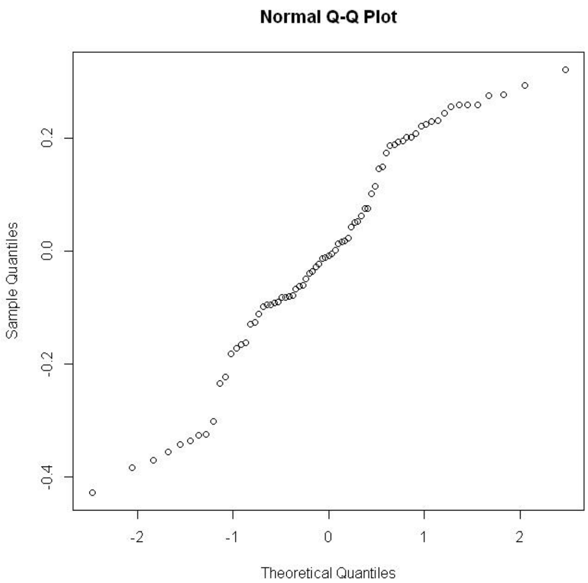
The formula is

$ln(Y_t) = \mu + e_t$

Where μ is the mean value of the log-transformation of the data.

Therefore, the formula can be written as $ln(Y_t) = 7.189148 + e_t$

We get the q-q plot for the residuals of the model as followed:



The Shapiro-Wilk normality test result shows that the statistic value is 0.9586 and the p-value is 0.01524. The p-value is lower than 0.05 and the residuals are not symmetrical, therefore, the model is rejected.

5. ARIMA(1,0,0) on the log-transformation of the gold price

The formula is

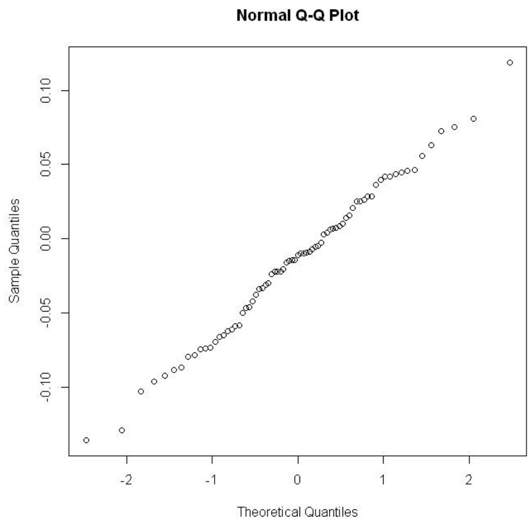
$\ln(Y_t) = \mu + \phi_1 ln(Y_{t-1}) + e_t$

Where μ is the mean value of the log-transformation data, ϕ_1 is the least squares estimation.

Therefore, the formula can be written as $ln(Y_t) = 7.189148 - 0.996 ln(Y_{t-1}) + e_t$

We get the q-q plot for the residuals of the model as followed:

The Shapiro-Wilk normality test result shows that the statistic value is 0.9934 and the p-value is 0.9679. The p-value and the q-q plot are satisfied the request and better than the third model. Next we will check the last model to find out if it is a better one.

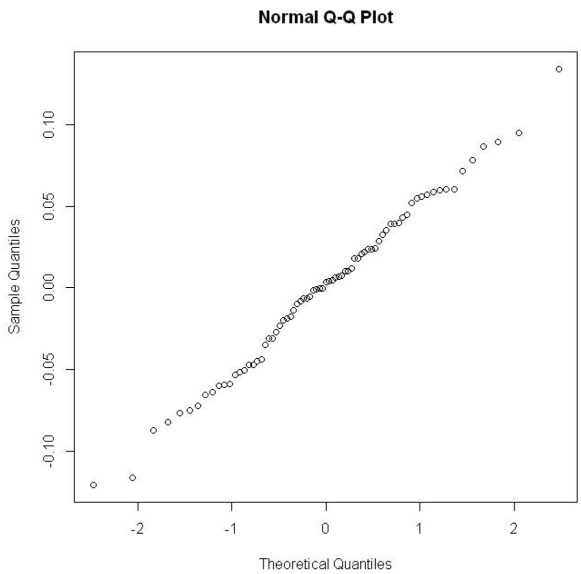


6. ARIMA(0,1,0) on the log-transformation of the gold price

The formula is

$W_t = \ln(Y_t) - \ln(Y_{t-1}) = e_t$

We get the q-q plot for the residuals of the model as followed:



The Shapiro-Wilk normality test result shows that the statistic value is 0.9932 and the p-value is 0.9627. The p-value and the q-q plot are satisfied the request but not better than the fifth model.

Conclusion

Six ARIMA (p, d, q) models were used to matching data. Based on q-q polt and Shapiro-Wilk normality test results, we found out that model 1 and model 4 are not satisfied with p-value lower than 0.05, therefore the two models are rejected. Model 2, 3, 5 and 6 are all satisfied the request. However, model 5 has the highest p-value and the most symmetrical residuals, therefore finally the model $ln(Y_t) = 7.189148 - 0.996ln(Y_{t-1}) + e_t$ was chosen.

Appendix.

No.	Gold Price	First difference of the gold price	Log-transformation of the gold price	First difference of Log-transformation
1	863.70	51.10	6.761225	0.057480
2	914.80	32.80	6.818705	0.035227
3	947.60	-18.20	6.853932	-0.019393
4	929.40	-25.50	6.834539	-0.027820
5	903.90	76.10	6.806719	0.080834
6	980.00	-38.70	6.887553	-0.040291
7	941.30	17.50	6.847262	0.018421
8	958.80	-2.30	6.865683	-0.002402
9	956.50	92.10	6.863281	0.091930
10	1048.60	11.30	6.955211	0.010719
11	1059.90	141.40	6.965930	0.125230
12	1201.30	-78.00	7.091160	-0.067134
13	1123.30	-19.20	7.024026	-0.017240
14	1104.10	12.20	7.006786	0.010989
15	1116.30	10.80	7.017775	0.009628
16	1127.10	40.70	7.027403	0.035474
17	1167.80	57.60	7.062877	0.048146
18	1225.40	-16.70	7.111023	-0.013722
19	1208.70	-23.10	7.097301	-0.019296
20	1185.60	62.70	7.078004	0.051534
21	1248.30	99.60	7.129538	0.076765
22	1347.90	8.10	7.206303	0.005991
23	1356.00	38.00	7.212294	0.027638
24	1394.00	-10.10	7.239933	-0.007272
25	1383.90	-42.00	7.232661	-0.030819
26	1341.90	87.80	7.201842	0.063378
27	1429.70	-0.50	7.265220	-0.000350
28	1429.20	103.30	7.264870	0.069786
29	1532.50	7.00	7.334656	0.004557
30	1539.50	-52.80	7.339213	-0.034899
31	1486.70	135.00	7.304314	0.086916
32	1621.70	207.40	7.391230	0.120349
33	1829.10	-158.30	7.511579	-0.090521
34	1670.80	41.00	7.421058	0.024243
35	1711.80	38.00	7.445301	0.021956
36	1749.80	-137.60	7.467257	-0.081902
37	1612.20	133.00	7.385355	0.079269
38	1745.20	-28.20	7.464624	-0.016291
39	1717.00	-86.10	7.448334	-0.051447
40	1630.90	23.40	7.396887	0.014246
41	1654.30	-32.20	7.411133	-0.019656

42	1622.10	-22.10	7.391477	-0.013718
43	1600.00	7.30	7.377759	0.004552
44	1607.30	85.00	7.382311	0.051533
45	1692.30	85.80	7.433844	0.049457
46	1778.10	-62.10	7.483301	-0.035549
47	1716.00	1.30	7.447751	0.000757
48	1717.30	-58.90	7.448509	-0.034900
49	1658.40	9.20	7.413609	0.005532
50	1667.60	-90.70	7.419141	-0.055925
51	1576.90	19.80	7.363216	0.012478
52	1596.70	-130.30	7.375694	-0.085129
53	1466.40	-55.60	7.290566	-0.038653
54	1410.80	-171.60	7.251912	-0.129691
55	1239.20	69.60	7.122221	0.054645
56	1308.80	87.00	7.176866	0.064357
57	1395.80	-76.10	7.241223	-0.056063
58	1319.70	-4.20	7.185160	-0.003188
59	1315.50	-93.60	7.181972	-0.073810
60	1221.90	0.50	7.108162	0.000409
61	1222.40	39.10	7.108571	0.031485
62	1261.50	88.80	7.140057	0.068025
63	1350.30	-71.00	7.208082	-0.054014
64	1279.30	30.20	7.154068	0.023332
65	1309.50	-64.50	7.177401	-0.050510
66	1245.00	82.10	7.126891	0.063861
67	1327.10	-31.90	7.190751	-0.024331
68	1295.20	-7.20	7.166420	-0.005574
69	1288.00	-66.10	7.160846	-0.052684
70	1221.90	-56.60	7.108162	-0.047428
71	1165.30	46.50	7.060734	0.039128
72	1211.80	-6.70	7.099862	-0.005544
73	1205.10	69.30	7.094318	0.055913
74	1274.40	-68.50	7.150231	-0.055249
75	1205.90	-2.30	7.094981	
76	1203.60			